LES P-VALUES COMME VOTES D'EXPERTS*

GUY MOREL[1]

ABSTRACT. The p-values are often implicitly used as a measure of evidence for the hypotheses of the tests. This practice has been analyzed with different approaches. It is generally accepted for the one-sided hypothesis problem, but it is often criticized for the two-sided hypothesis problem. We analyze this practice with a new approach to statistical inference. First we select good decision rules without using a loss function, we call them experts. Then we define a probability distribution on the space of experts. The measure of evidence for a hypothesis is the inductive probability of experts that decide this hypothesis.

RÉSUMÉ. Dans la pratique des tests, les p-values sont souvent utilisées comme des mesures du niveau de confiance à accorder aux hypothèses. Analysée de différents points de vue, cette manière de faire est généralement acceptée pour les hypothèses unilatérales. Elle est par contre souvent critiquée dans le cas d'hypothèses bilatérales. Nous reprenons ce débat en utilisant une approche nouvelle des problèmes de décision. Elle consiste à sélectionner de bonnes règles de décision, appelées experts, sans utiliser de fonction de perte. L'espace des experts est ensuite probabilisé. Le poids des experts qui décident une hypothèse est pris comme indice de confiance en faveur de cette hypothèse.

AMS Subject Classification.. 62A99, 62C05, 62P.

1. Introduction

Tous les logiciels statistiques fournissent le résultat d'un test en donnant la p-value (le niveau de signification). Ce seuil minimum de rejet permet de conclure quel que soit le seuil choisi. Si la suite du programme ne dépend pas du résultat du test, l'utilisateur n'est pas obligé de fournir un seuil. On lui évite ainsi bien des tâtonnements, car dans beaucoup d'applications le seuil n'est pas complètement fixé. Plusieurs valeurs standards sont envisageables, par exemple $\alpha = 0.05$, $\alpha = 0.01$ et $\alpha = 0.001$. Lorsque le rejet de l'hypothèse H_0 est possible, le résultat est souvent symbolisé par des étoiles : * signifie que seul $\alpha = 0.05$ permet de rejeter H_0, *** indique que même $\alpha = 0.001$ le permet. Cette symbolique des étoiles n'est bien sûr pas innocente, elle

URL address of the journal: **http://www.emath.fr/ps/**

Mots clés. Théorie de la décision, tests, p-values, seuils minimum de rejet, hypothèses unilatérales et bilatérales.

* Recherche réalisée dans le cadre du LAST et du CNRS UPRES-A 6083 de Tours.

[1] Université de Tours, UFR Arts et Sciences Humaines, 3 Rue des Tanneurs, 37041 Tours Cedex, France ; e-mail : `morel@univ-tours.fr`

traduit que le rejet est d'autant plus sûr que les étoiles sont nombreuses. Donner la p-value quand elle est inférieure à 0.05 se fait de plus en plus. Ce n'était pas possible lorsque le résultat d'un test s'obtenait par lecture sur une table, l'utilisateur ne disposait que des valeurs critiques du test pour les seuils tabulés. Avec la p-value la symbolique des étoiles se transforme en une phrase du type : plus la p-value est faible, plus la "probabilité" de l'hypothèse H_0 est faible. Souvent cette interprétation probabiliste s'élargit à toute la plage des valeurs possibles, l'intervalle $[0, 1]$. La p-value est alors considérée comme la "probabilité" de l'hypothèse H_0, son complément à 1 comme la "probabilité" de H_1. Les guillemets posent le problème du statut de cette probabilité, les auteurs de la théorie des tests n'avaient pas pour but de fournir une probabilité pour les deux hypothèses. Les p-values sont un indice du niveau de confiance accordé à l'hypothèse H_0 qui a été forgé par la pratique des tests. Cet indice statistique non théorisé suscite de nombreux articles. On peut les diviser en deux grandes catégories.

Certains articles partent de l'outil p-value tel qu'il est construit par la pratique et étudient ses propriétés. Parmi les plus récents, citons par exemple l'article de Hung, O'Neill, Bauer et Köhne (1997) qui traite de la distribution des p-values sous l'hypothèse H_1, et celui de Thompson (1996) qui utilise les p-values, à la place de la puissance, pour comparer les tests.

Les autres articles traitent plutôt du fondement des p-values comme probabilité sur l'espace des deux hypothèses. Ils les étudient dans un autre cadre que la théorie des tests. C'est aussi ce que nous essaierons de faire ici.
Le cadre théorique qui paraît le plus approprié pour ce type d'étude est le cadre bayésien : les p-values peuvent-elles être considérées comme des probabilités a posteriori ? Pour résumer, disons que ce n'est généralement pas le cas, mais que pour les tests unilatéraux les p-values s'obtiennent souvent à partir de probabilités a priori non informatives [Berger (1985) ; Berger et Sellke (1987) ; Casella et Berger (1987) ; Robert (1992)].

On peut aussi étudier les p-values comme estimateurs de l'indicatrice de l'ensemble des paramètres définissant H_0 [Hwang, Casella, Robert, Wells et Farrell (1992) ; van der Meulen et Schaafsma (1993) ; Robert (1992) ; Schaafsma, Tolboom et van der Meulen (1989)]. Là encore, pour les tests unilatéraux les p-values sont généralement admissibles, alors qu'elles ne le sont pas pour les tests bilatéraux.

Cette opposition entre hypothèses unilatérales et hypothèses bilatérales se retrouve quand on regarde la cohérence des p-values lorsqu'on fait varier l'hypothèse H_0, sans changer le modèle statistique. Considérons deux hypothèses possibles $H_0 : \theta \in \Theta_0$ $(\Theta = \Theta_0 + \Theta_1)$ et $H_0' : \theta \in \Theta_0'$ $(\Theta = \Theta_0' + \Theta_1')$. Si $\Theta_0 \subset \Theta_0'$ on cherche généralement une mesure du niveau de confiance qui donne une valeur plus grande pour Θ_0' que pour Θ_0. Ce n'est pas le cas des p-values pour des hypothèses bilatérales, Schervish (1996) le montre dans le modèle statistique défini par la famille des lois gaussiennes $\{N(\theta, 1)\}_{\theta \in \mathbb{R}}$. Il trouve facilement $\theta_1 < -\theta_0 < \theta_0 < \theta_2 < x$, tels que les hypothèses $[\theta_0] \subset [-\theta_0, \theta_0] \subset [\theta_1, \theta_2]$ et la réalisation x donnent des p-values qui sont décroissantes au lieu d'être croissantes. Par exemple, pour $\theta_1 = -0.82$, $\theta_0 = 0.50$, $\theta_2 = 0.52$ et $x = 2.18$ il obtient les p-values suivantes : 0.0930 si $\Theta_0 = [\theta_0]$, 0.0502 si $\Theta_0 = [-\theta_0, +\theta_0]$ et 0.0498 si $\Theta_0 = [\theta_1, \theta_2]$. Ce sont les

valeurs respectives de $2[1-\Phi(2.18-0.50)]$, $[1-\Phi(2.18-0.50)]+[1-\Phi(2.18+0.50)]$ et $[1-\Phi(2.18-0.52)]+[1-\Phi(2.18+0.82)]$, Φ étant la fonction de répartition de la loi normale réduite (voir le paragraphe 4.2).

Les p-values peuvent aussi être retrouvées à partir de la distribution fiduciaire de Fisher mais ce dernier tenait à bien distinguer les deux concepts [Salomé (1998) p. 67 et p. 89]. L'inférence fiduciaire sert rarement de point de vue pour juger les p-values, peut-être parce qu'il n'existe pas une définition suffisamment générale [Buehler (1980)].

Notre regard sur les p-values provient d'une manière différente de poser le problème du choix entre deux hypothèses. Nous ne cherchons pas à sélectionner une règle de décision, nous nous contentons d'un critère définissant les "bonnes" règles, celles que nous appellerons experts. Les décisions prises par ces experts sont ensuite synthétisées dans un vote qui donne un indice de confiance pour chacune des deux hypothèses. Dans un modèle à rapport de vraisemblance monotone, la p-value d'un test unilatéral est un vote particulier. Pour les tests bilatéraux ceci n'est généralement possible qu'en changeant le cadre décisionnel, plus précisément la tribu du modèle statistique. Avant de préciser ces résultats il nous faut définir les notions d'expert et de vote.

2. Expertises

L'étude d'un problème de décision statistique passe par la donnée de critères de sélection entre les différentes règles de décision considérées. Classiquement on commence par se donner une fonction de perte et on compare les procédures de décision à partir des fonctions de risque correspondantes. Il est rare qu'un choix unique de ce critère s'impose, bien que l'étude de certaines fonctions de perte soit privilégiée, par exemple la perte quadratique dans le cadre de l'estimation ou le risque de se tromper pour le choix entre deux hypothèses. Même si ce dernier choix paraît assez "naturel", d'autres pertes sont possibles, en particulier si on regarde ce problème de choix entre deux hypothèses comme un problème d'estimation [Hwang, Casella, Robert, Wells et Farrell (1992) ; Robert (1992)]. Les fonctions de risque permettent d'introduire un préordre partiel sur les règles de décision et ainsi de sélectionner les règles admissibles. Cette première sélection s'obtient en comparant les règles entre elles, elle ne dérive pas d'une propriété intrinsèque. Nous avons cherché à imposer ce type de propriété pour définir les "bonnes" règles de décision, celles que nous appellerons experts. Une règle pourra être déclarée expert sans avoir à la comparer à l'ensemble des autres règles. La propriété de base que nous imposons aux experts est simplement que le diagnostic d, θ appartient à Θ_d, doit être plus probable quand θ appartient à Θ_d que lorsque θ n'y appartient pas. C'est une notion de règle non biaisée, mais pour que cela suppose une connaissance fine du modèle nous imposons que cette propriété soit aussi vérifiée conditionnellement à tout événement non négligeable, et pas simplement globalement comme c'est généralement le cas. Dans le cadre de cet article, nous ne donnerons la définition des experts que pour les modèles statistiques $(\Omega, \mathcal{A}, (\mathbb{P}_\theta = p_\theta.\mu)_{\theta\in\Theta})$ ayant des densités

strictement positives. Ceci évite les complications liées au glissement des supports des probabilités $\mathbb{P}_\theta$ (le cas général est traité dans Morel(1997)).

Définition 2.1.
Considérons le problème de décision défini par la partition : $\Theta = \Theta_0 + \Theta_1$. *La règle de décision* $\phi : (\Omega, \mathcal{A}) \to \{0,1\}$ *est un expert du choix entre* Θ_0 *et* Θ_1 *si pour tout couple* $(\theta_0, \theta_1) \in \Theta_0 \times \Theta_1$ *et tout événement non négligeable* $C \in \mathcal{A}$ *elle vérifie :*
$\mathbb{P}_{\theta_1}(C \cap \{\phi = 1\})/\mathbb{P}_{\theta_1}(C) \geq \mathbb{P}_{\theta_0}(C \cap \{\phi = 1\})/\mathbb{P}_{\theta_0}(C)$.

Cette propriété est intrinsèque, elle ne fait pas intervenir d'autres règles que ϕ. Dans le contexte d'une réflexion sur les p-values on pourrait juger pertinent de considérer les règles de décision à valeurs dans $[0,1]$. On rajouterait ainsi de nombreuses règles, par exemple toutes les fonctions croissantes du rapport $p_{\theta_1}/p_{\theta_0}$ pour le choix entre deux probabilités. Il deviendrait alors difficile de définir une probabilité sur l'ensemble des experts (voir le paragraphe 2-1). Nous pouvons justifier notre démarche en disant qu'au départ, on cherche à définir des experts qui prennent totalement leurs responsabilités en choisissant l'une des deux hypothèses. La recherche d'un indice de confiance pour les hypothèses se fait dans une deuxième étape, si différents experts sont possibles et si l'on ne tient pas à essayer d'en choisir un.

Un expert du choix entre Θ_0 et Θ_1 est aussi un expert de tout problème de décision emboîté défini par : $\Theta' = \Theta'_0 + \Theta'_1$, $\Theta'_0 \subseteq \Theta_0$ et $\Theta'_1 \subseteq \Theta_1$. Le passage d'un modèle emboîté au modèle de départ ne peut que réduire l'ensemble des experts. Ceux ci peuvent donc se restreindre à $1\!\!\mathrm{I}_\emptyset$ et $1\!\!\mathrm{I}_\Omega$. Nous verrons que c'est le cas des problèmes bilatères. Le travail sur les experts est ainsi différent de celui fait sur les règles admissibles dans la théorie de la décision à partir d'une fonction de perte. En effet, les règles admissibles pour un modèle emboîté sont généralement encore admissibles dans le modèle de départ. Les règles admissibles sont souvent trop nombreuses, les experts eux sont plutôt trop rares. Ceci nous amène à distinguer deux cas.

2.1. Le problème de décision est expertisable

Dans ce cas l'ensemble des experts ne se réduit pas aux experts triviaux. Pour cela il faut que les problèmes de décision emboîtés soient semblables. Par exemple dans la théorie des tests, les propriétés obtenues avec des hypothèses simples (lemme de Neyman-Pearson principalement) se prolongent facilement au cas des tests unilatéraux dans un modèle à rapport de vraisemblance monotone. Ce qui est important dans ce type de problème de décision, c'est que le problème du choix entre un élément θ_0 de Θ_0 et un élément θ_1 de Θ_1 ne change pas fondamentalement lorsque θ_0 et θ_1 varient. Il existe une statistique réelle T qui ordonne les observations, des plus favorables à θ_1 aux plus favorables à θ_0. Les bonnes règles de décision sont de la forme $1\!\!\mathrm{I}_{\{T<t\}}$ et $1\!\!\mathrm{I}_{\{T\leq t\}}$, nous verrons qu'il en est de même des experts. Face à ce trop plein de bonnes règles déterministes le statisticien se donne généralement des critères supplémentaires pour essayer de sélectionner une règle de décision : recherche d'un test U.P.P., d'un test sans biais U.P.P., d'un test invariant U.P.P., d'une règle de Bayes, d'une règle minimax, etc.

Ceci est nécessaire lorsque l'utilisateur cherche à mettre au point une règle de décision automatique. Dans d'autres contextes c'est plutôt une aide à la décision qui est demandée, un indice de confiance pour chacune des hypothèses. Cette perspective nous conduit à essayer de résumer les choix de l'ensemble des experts. Nous allons considérer nos experts comme égaux en droit, les faire voter et fournir à l'utilisateur le résultat de ce vote.
Pour faire voter nos experts il faut définir une probabilité sur l'ensemble des experts. Nous en avons défini une à partir de chaque probabilité $\mathbb{P}_\theta$ en essayant d'accorder d'autant plus de poids à un ensemble d'experts qu'il est formé d'experts donnant des résultats différents sous $\mathbb{P}_\theta$. Par exemple, les experts $\{1\!\!\mathrm{I}_{\{T<t\}}, 1\!\!\mathrm{I}_{\{T\leq t\}}; t' < t < t''\}$ donnent la même décision sur $\{T \leq t'\}$ et $\{T \geq t''\}$, ils auront d'autant plus de poids dans le vote que $\mathbb{P}_\theta(\{t' < T < t''\})$ sera grand. Il est alors impossible de "bourrer les urnes" avec des experts presque sûrement égaux. Nous avons bien sûr autant de votes que de valeurs du paramètre et il y a bien des manières de les synthétiser. On peut faire une moyenne à partir d'une probabilité sur l'ensemble des paramètres en prenant en compte des informations a priori. On peut aussi sélectionner le vote le plus neutre par rapport aux deux hypothèses. C'est ce type de solution que nous étudierons ici, car il permet de donner un sens nouveau à la notion de p-value (la prise en compte d'informations a priori est traitée dans Morel (1997)).

2.2. Le problème de décision n'est pas expertisable

Dans ce cas il n'y a que les experts triviaux. C'est ce qui se passe généralement pour les hypothèses bilatérales $H : \theta \in]-\infty, \theta_1[\cup]\theta_2, +\infty[$ et $H' : \theta \in [\theta_1, \theta_2]$. Pour obtenir un problème expertisable il faut diminuer les contraintes imposées par la définition des experts en travaillant sur une sous-tribu. Ceci diminue l'ensemble des événements sur lesquels la propriété de base des experts doit s'appliquer conditionnellement. Notons H_- et H_+ les deux parties de l'hypothèse englobante H. Si l'on ne veut pas les différencier, il ne faut pas dissocier les observations qui sont en faveur de H_- de celles qui, avec la même "force", sont en faveur de H_+. Ceci revient à ne considérer que les événements fournissant une information symétrique sur H_- et H_+. On retrouvera alors comme vote les p-values des tests sans biais.
Cette solution ne nous semble pas la plus intéressante. Dans bien des cas, le choix principal est entre H et H' mais H_- et H_+ ne signifient pas la même chose. Les tests bilatéraux sont souvent utilisés comme s'ils étaient construits pour choisir entre les trois hypothèses H_-, H' et H_+. Dans ce cas il est alors intéressant que le vote par rapport à H et H' soit cohérent avec les deux votes correspondant aux hypothèses unilatérales $\{H_-; H' \cup H_+\}$ et $\{H_- \cup H'; H_+\}$ [Gabriel (1969)]. Le poids des votes en faveur de H doit être supérieur à celui des votes en faveur de H_- (resp. H_+). Nous avons pour cela introduit la notion de votes compatibles sur une famille d'hypothèses unilatérales. Cette manière de faire a l'avantage d'imposer une cohérence entre les solutions de problèmes de décision qui reposent sur une même structuration de l'espace des paramètres.

Nous allons maintenant faire fonctionner cette notion d'expertise sur quelques problèmes de décision classiques.

3. Hypothèses unilatérales dans un modèle à rapport de vraisemblance monotone

Dans la théorie des tests la première application du lemme de Neyman-Pearson concerne les hypothèses unilatérales dans un modèle à rapport de vraisemblance monotone. Le cas le plus couramment traité est celui d'une statistique réelle exhaustive X de loi $\mathbb{P}_\theta = p_\theta.\mu$, le paramètre θ appartenant à un intervalle Θ de $\mathbb{R}$, la propriété de vraisemblance monotone exprimant la croissance stricte du rapport $p_{\theta''}/p_{\theta'}$ lorsque $\theta' < \theta''$ [Monfort (1982)]. Nous nous restreindrons ici à des densités par rapport à la mesure de Lebesgue λ, continues en θ. Dans les autres modèles les conclusions restent identiques mais nous éviterons les problèmes techniques induits par la présence de masses ou par la discontinuité de la famille $\{p_\theta\}_{\theta\in\Theta}$ (le cas général est traité dans Morel (1997 ou 1998)). Ce cas particulier recouvre un grand nombre de cas classiques : les modèles exponentiels à paramètre réel mais aussi les problèmes statistiques portant sur le paramètre de non centralité d'une famille de densités de Student, Fisher ou khi-deux [Karlin (1955)].

Soit $\theta_1 \in \Theta$, nous allons étudier le problème du choix entre $\Theta_1 =]-\infty, \theta_1]$ et $\Theta_0 =]\theta_1, +\infty[$ (grâce à la continuité des p_θ en θ_1, prendre $\Theta_1 =]-\infty, \theta_1[$ et $\Theta_0 = [\theta_1, +\infty[$ ne change rien). Notons $F(\theta, x)$ la valeur en x de la fonction de répartition de X pour la probabilité $\mathbb{P}_\theta$.
Dans la théorie des tests il y a deux traitements distincts suivant que l'on choisit de privilégier Θ_0 ou Θ_1. Le test de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ au seuil α, rejette H_0 pour $x < t_\alpha$ avec $F(\theta_1, t_\alpha) = \alpha$. Le test de $H_0' : \theta \in \Theta_1$ contre $H_1' : \theta \in \Theta_0$ au seuil α, rejette H_0' pour $x > t'_\alpha$ avec $F(\theta_1, t'_\alpha) = 1 - \alpha$. Lorsqu'on réalise x, les logiciels statistiques donnent pour le premier test la p-value $\alpha_{\theta_1}(x) = F(\theta_1, x)$ et pour le deuxième test la p-value $\alpha'_{\theta_1}(x) = 1 - F(\theta_1, x)$. Ces seuils minimum de rejet $\alpha_{\theta_1}(x)$ et $\alpha'_{\theta_1}(x) = 1 - \alpha_{\theta_1}(x)$ sont interprétés comme des indices statistiques de la vraisemblance de chacune des hypothèses Θ_0 et Θ_1. Les p-values des tests unilatéraux sont bien cohérentes, pour $\theta_1 < \theta_2$ on a $\Theta_0 =]\theta_1, +\infty[\supset]\theta_2, +\infty[$ et $\alpha_{\theta_1}(x) > \alpha_{\theta_2}(x)$.
Nous allons retrouver ces p-values comme votes d'experts. Les experts du choix entre Θ_1 et Θ_0 sont les différents tests possibles entre ces deux hypothèses.

Proposition 3.1.

Sur un intervalle $I \subseteq \mathbb{R}$ muni de la mesure de Lebesgue, considérons une famille $\{p_\theta\}_{\theta\in\Theta}$ de densités strictement positives. Θ étant un intervalle de $\mathbb{R}$, cette famille est supposée à rapport de vraisemblance strictement monotone. Les experts du choix entre les deux hypothèses unilatérales $\Theta_1 =]-\infty, \theta_1] \cap \Theta \neq \emptyset$ et $\Theta_0 =]\theta_1, +\infty[\cap \Theta \neq \emptyset$ sont les règles de décision à valeurs dans $\{0, 1\}$, presque sûrement de la forme $f_t(x) = \mathbb{1}_{]-\infty, t[}(x)$ avec $t \in \overline{\mathbb{R}}$.

Démonstration.

I – $f_t = 1\!\!1_{]-\infty,t[}$ est un expert.

Les événements non négligeables sont les boréliens dont l'intersection avec l'intervalle I est de mesure de Lebesgue non nulle. Soient C un de ces événements, $\theta' \leq \theta_1$ et $\theta'' > \theta_1$. On doit démontrer l'inégalité :
$\mathbb{P}_{\theta'}(C \cap]-\infty, t[)/\mathbb{P}_{\theta'}(C) \geq \mathbb{P}_{\theta''}(C \cap]-\infty, t[)/\mathbb{P}_{\theta''}(C)$.

On va travailler avec la famille des probabilités conditionnelles $\mathbb{P}_\theta^C$ de densité $(1/\mathbb{P}_\theta(C))\ 1\!\!1_C$ par rapport à $\mathbb{P}_\theta$. f_t est un test de Neyman et Pearson pour tester $\mathbb{P}_{\theta''}^C$ contre $\mathbb{P}_{\theta'}^C$ au seuil $\mathbb{P}_{\theta''}^C(]-\infty,t[)$. La puissance étant supérieure au seuil [Lehmann (1986) p. 76], on a $\mathbb{P}_{\theta'}^C(]-\infty,t[) \geq \mathbb{P}_{\theta''}^C(]-\infty,t[)$, ce qui est bien l'inégalité recherchée.

II – Un expert ϕ est presque sûrement de la forme f_t.

Posons $t' = sup\{t \in \overline{\mathbb{R}}\,;\, f_t \overset{p.s.}{\leq} \phi\}$ et $t'' = inf\{t \in \overline{\mathbb{R}}\,;\, f_t \overset{p.s.}{\geq} \phi\}$. On doit démontrer que le cas $t' < t''$ est impossible lorsque ϕ est un expert. Pour cela nous allons supposer $t' < t''$ et trouver un événement C non négligeable vérifiant : $\mathbb{P}_{\theta'}(C \cap \{\phi = 1\})/\mathbb{P}_{\theta'}(C) < \mathbb{P}_{\theta''}(C \cap \{\phi = 1\})/\mathbb{P}_{\theta''}(C)$ lorsque $\theta' \leq \theta_1 < \theta''$.
Soit $t \in]t', t''[$.
Considérons l'événement $A = [t', t[\cap \{\phi = 0\}$, par définition de t' il n'est pas Lebesgue négligeable. De même, la définition de t'' entraîne que l'événement $B =]t, t''] \cap \{\phi = 1\}$ n'est pas négligeable.
Posons $C = A \cup B$, on a $\mathbb{P}_\theta(C \cap \{\phi = 1\})/\mathbb{P}_\theta(C) = \mathbb{P}_\theta(B)/[\mathbb{P}_\theta(A) + \mathbb{P}_\theta(B)]$. Il est facile de voir que l'événement C vérifie l'inégalité recherchée si et seulement si $\mathbb{P}_{\theta'}(A)/\mathbb{P}_{\theta'}(B) > \mathbb{P}_{\theta''}(A)/\mathbb{P}_{\theta''}(B)$. Cette dernière inégalité est une conséquence directe de la propriété de vraisemblance strictement monotone ; notons k le rapport $p_{\theta''}(t)/p_{\theta'}(t)$, pour $x \in A$ on a : $p_{\theta''}(x) < kp_{\theta'}(x)$ et pour $x \in B$: $p_{\theta''}(x) > kp_{\theta'}(x)$, donc $\mathbb{P}_{\theta''}(A) < k\mathbb{P}_{\theta'}(A)$ et $\mathbb{P}_{\theta''}(B) > k\mathbb{P}_{\theta'}(B)$. $\square$

Chacun de ces experts traduit un a priori plus ou moins fort en faveur de l'une des deux hypothèses. Plus t est grand et plus l'expert f_t parie sur l'hypothèse $\theta \in \Theta_1$. Si l'on voulait en choisir un il faudrait imposer des contraintes supplémentaires. Dans la théorie des tests, le choix de l'hypothèse qui jouera le rôle de H_0 revient à privilégier les experts qui parient en faveur de H_0. Nous avons préféré garder l'ensemble des avis des experts et essayer de les résumer. Se pose alors le problème du poids à accorder à chacun de ces avis, il faut probabiliser l'ensemble des experts : $\{f_t\}_{t \in \overline{\mathbb{R}}}$. Nous quittons ici une approche de type Neyman-Wald. La difficulté du choix des critères de sélection est remplacée par celle de la définition du vote des experts.

Pour probabiliser l'espace des experts nous le munissons de la tribu définie à partir de l'ordre primordial introduit par l'indice $t \in \overline{\mathbb{R}}$. Nous donnerons à un intervalle d'experts $[f_{t'}, f_{t''}] = \{f_t\,;\, t' \leq t \leq t''\}$ d'autant plus de poids qu'il est formé d'experts très différents. Si θ est la vraie valeur du paramètre il est assez naturel de juger les différences entre experts en utilisant la probabilité $\mathbb{P}_\theta$. Les experts $[f_{t'}, f_{t''}]$ sont égaux en dehors de l'intervalle $[t', t''[$ nous leur accorderons d'autant plus d'importance que la probabilité $\mathbb{P}_\theta([t', t''[)$ est grande. Pour chaque $\theta \in \mathbb{R}$ nous pouvons ainsi définir sur l'ensemble des experts une probabilité inductive dont la fonction de répartition est

$F(\theta,.)$. Lorsqu'on réalise x, les experts qui décident $\theta \in \Theta_1$ sont ceux correspondant à $t > x$. Cet ensemble d'experts $\{f_t\,;\, t > x\}$ a alors un poids égal à $1 - F(\theta, x)$. Ceci nous conduit à la notion de vote.

Définition 3.2.

Considérons le problème du choix entre les deux hypothèses unilatérales $\Theta_1 =]-\infty, \theta_1] \cap \Theta \neq \emptyset$ et $\Theta_0 =]\theta_1, +\infty[\cap \Theta \neq \emptyset$, dans le modèle de la proposition 3-1. On appelle vote des experts sous $\mathbb{P}_\theta$ la famille de probabilités inductives $\{Q^x_\theta\}_{x\in\mathbb{R}}$ définie sur l'espace $\{0,1\}$, correspondant aux deux hypothèses, par $Q^x_\theta(\{1\}) = 1 - F(\theta, x)$.

Nous avons ainsi autant de votes qu'il y a de valeurs du paramètre. Il nous faut maintenant choisir un de ces votes ou les résumer en définissant un mélange. Pour cela nous pouvons faire intervenir une information a priori. Les exemples d'application où une telle information existe sont cependant rares. L'analyse bayésienne, qui repose sur la donnée d'une probabilité a priori, est appliquée la plupart du temps sans utiliser d'information a priori, en partant par exemple d'une loi a priori non informative. C'est d'ailleurs ce type de lois a priori qui permet de retrouver certaines p-values comme probabilités a posteriori. Nous allons obtenir ces p-values comme vote d'experts en cherchant un vote ne reposant sur aucune information a priori. Nous n'étudierons donc pas ici la possibilité de faire intervenir ce type d'information (ce cas est traité dans Morel (1998)).
Le vote des experts sous $\mathbb{P}_\theta$ est d'autant plus favorable à la décision $d = 1$ que θ est grand. Sous Θ_0 on ne veut pas trop favoriser la décision $d = 1$ et inversement sous Θ_1 on ne veut pas trop favoriser la décision $d = 0$. Pour chacune de ces hypothèses on considère le vote qui lui est le plus favorable. Sous Θ_1 c'est le vote défini par le point frontière θ_1. Il en est de même sous Θ_0 car la famille $\{p_\theta\}$ est continue en θ. Le vote correspondant à $\theta = \theta_1$ a la propriété d'être neutre, de n'avantager aucune des deux hypothèses au sens suivant : $\mathbb{E}_{\theta_1}[Q^x_{\theta_1}(\{1\})] = \mathbb{E}_{\theta_1}[Q^x_{\theta_1}(\{0\})] = 1/2$; c'est une conséquence directe du fait que les valeurs de la fonction de répartition se répartissent uniformément sur $[0, 1]$.
Pour une réalisation x, le vote au point frontière (sous $\mathbb{P}_{\theta_1}$) définit une probabilité inductive $Q^x_{\theta_1}$ qui donne les p-values : $Q^x_{\theta_1}(\{1\}) = 1-F(\theta_1, x) = \alpha'_{\theta_1}(x)$ et $Q^x_{\theta_1}(\{0\}) = \alpha_{\theta_1}(x)$. Ceci fournit une légitimité supplémentaire à l'interprétation de ces p-values comme indice de vraisemblance des hypothèses unilatérales $\theta \le \theta_1$ et $\theta > \theta_1$.

Exemple 3.3.

En analyse de variance à effets fixes on utilise des statistiques qui suivent des lois de Fisher décentrées [Scheffé(1970) p. 38]. Le paramètre de non centralité λ étant la quantité, qui intéresse l'utilisateur, exprimée par rapport à l'écart-type commun des variables. Par exemple, dans une analyse de variance classique à deux facteurs, λ est égal à $\sqrt{\sum(\theta_{ij})^2}/\sigma$ pour le test sur l'additivité des facteurs, et égal à $\sqrt{\sum(\alpha_i)^2}/\sigma$ pour le test sur la nullité des effets additifs du premier facteur. Si une statistique W suit une loi de Fisher décentrée de paramètre λ et de degrés de liberté (k, l), nous savons que $(k/l)W$ suit une loi bêta sur $\mathbb{R}^+$: $\beta(\frac{k}{2}, \frac{l}{2}, \frac{\lambda^2}{2})$ [Barra (1971) p. 84]. Nous

allons étudier ce modèle statistique paramétré par $\theta = \frac{\lambda^2}{2}$.

Considérons sur $\mathbb{R}^+$ la famille des lois bêta décentrées : $\beta(p, q, \theta)$, les paramètres $p > 0$ et $q > 0$ sont connus, le paramètre de non centralité $\theta \geq 0$, lui, est inconnu. Comme pour la famille des lois de Fisher décentrées, on montre qu'elle est à rapport de vraisemblance strictement monotone [Karlin (1955)]. Soient $\theta_1 \geq 0$ et une réalisation $x \in \mathbb{R}^+$, le vote neutre en faveur de $\Theta_1 = [0, \theta_1]$ est égal à :

$$\begin{aligned} Q^x_{\theta_1}(\{1\}) &= 1 - F(\theta_1, x) = \int_x^{+\infty} [\sum_{m=0}^{+\infty} e^{-\theta_1} . \frac{\theta_1^m}{m!} \frac{\Gamma(p+m+q)}{\Gamma(p+m)\Gamma(q)} \frac{y^{p+m-1}}{(1+y)^{p+m+q}}]\, dy \\ &= \int_{\mathbb{N}} [1 - F(p+m, q, x)]\, d\mathcal{P}_{\theta_1}(m), \end{aligned}$$

$\mathcal{P}_{\theta_1}$ étant une loi de Poisson de paramètre $\theta_1 \geq 0$ et $F(p+m, q, x)$ la valeur en x de la fonction de répartition d'une loi $\beta(p+m, q)$. Pour $\theta_1 = 0$ on obtient $1 - F(p, q, x)$, c'est le seuil minimum de rejet du test de $H_0 : \theta = 0$ contre $H_1 : \theta > 0$ (test classique en analyse de variance).

La notion d'expertise apporte une justification aux p-values de l'analyse de variance. Elle permet aussi de traiter des hypothèses non nulles de la forme $\theta \in [0, \theta_1]$, le résultat étant celui donné par l'inférence fiduciaire.

4. Hypothèses bilatérales

Nous allons étudier le problème du choix entre $\Theta_1 =]-\infty, \theta_1[\cup]\theta_2, +\infty[$ et $\Theta_0 = [\theta_1, \theta_2]$ avec $\theta_1 < \theta_2$ lorsque le modèle statistique vérifie les propriétés imposées dans le paragraphe précédent (prendre $\Theta_0 =]\theta_1, \theta_2[$ ne change rien, car les densités p_θ sont supposées continues en θ).

Dans la théorie des tests il y a encore deux traitements distincts suivant que l'on choisit de privilégier Θ_0 ou Θ_1. Pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ il n'existe pas de test uniformément plus puissant, on se restreint aux tests sans biais. L'étude de ces tests est classiquement faite dans les modèles exponentiels sur $\mathbb{R}$ de la forme $dP_\theta(\omega) = c(\theta) exp(\theta X(\omega)).d\mu(\omega)$ [Lehmann (1986), Monfort (1982)]. Dans le cadre de cet article, seul le cas où la mesure μ est diffuse nous intéresse. Au seuil α le plus puissant des tests sans biais rejette alors H_0 pour $x < c_1$ et $x > c_2$, c_1 et c_2 étant déterminés par $F(\theta_1, c_1) + [1 - F(\theta_1, c_2)] = F(\theta_2, c_1) + [1 - F(\theta_2, c_2)] = \alpha$. Pour tester $H'_0 : \theta \in \Theta_1$ contre $H'_1 : \theta \in \Theta_0$ au seuil α il existe un test uniformément plus puissant. Il rejette H'_0 lorsque $c'_1 \leq x \leq c'_2$, c'_1 et c'_2 étant déterminés par $F(\theta_1, c_2) - F(\theta_1, c_1) = F(\theta_2, c_2) - F(\theta_2, c_1) = \alpha$.

Les seuils minimum de rejet fournis par ces tests sont souvent interprétés comme des indices statistiques de la vraisemblance de chacune des hypothèses Θ_0 et Θ_1. Nous avons vu dans l'introduction que cette manière de faire est soumise à de nombreuses critiques. Nous allons montrer que dans le cadre d'une expertise elle n'est justifiée qu'en changeant le modèle statistique.

Tel qu'il est posé, le problème de décision entre deux hypothèses bilatérales ne possède que $\mathbb{1}_\emptyset$ et $\mathbb{1}_\mathbb{R}$ comme experts. Ceci est du au fait que tout expert du choix entre deux hypothèses est aussi un expert des problèmes de décision emboîtés, en particulier ici du choix entre les hypothèses unilatérales $]-\infty, \theta_1[$ et $\Theta_0 = [\theta_1, \theta_2]$, mais aussi du choix entre Θ_0 et $]\theta_2, +\infty[$. C'est

le même argument qui montre qu'il n'y a pas de test uniformément plus puissant de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$.
Face à cette pénurie d'experts nous envisagerons deux traitements possibles. On peut vouloir une cohérence entre la solution à ce problème et les expertises des hypothèses unilatérales dérivées. Cette manière de faire permet de conserver le sens donné par l'ordre qui structure l'espace des paramètres. On peut aussi envisager de casser cet ordre en créant une équivalence entre les éléments des deux parties de Θ_1 : $\Theta_1^- =]-\infty, \theta_1[$ et $\Theta_1^+ =]\theta_2, +\infty[$, c'est ce que fait la contrainte de non-biais imposée dans la théorie des tests de Θ_0 contre Θ_1.

4.1. Expertises compatibles

Bien souvent le choix entre Θ_1 et Θ_0 cache un problème de décision entre trois hypothèses : Θ_1^-, Θ_0 et Θ_1^+. L'utilisateur veut d'abord savoir si θ appartient à Θ_1 ou Θ_0, mais le regroupement de Θ_1^- et Θ_1^+ peut cacher des interprétations de type différent suivant que θ est inférieur à θ_1 ou supérieur à θ_2. Par exemple, si $\Theta_0 = [-\theta_0, +\theta_0]$ est la traduction d'un effet négligeable, le rejet de cette hypothèse conduit généralement dans la pratique à des conclusions différentes suivant que l'estimation de θ est positive ou négative. Dans un tel cadre le choix entre les hypothèses unilatérales Θ_1^- et $\Theta_0 \cup \Theta_1^+$ est envisageable, ainsi que le choix entre $\Theta_1^- \cup \Theta_0$ et Θ_1^+. Il est alors primordial que les trois problèmes de décision générés par Θ_1^-, Θ_0 et Θ_1^+ aient des solutions cohérentes. Les votes sélectionnés dans les deux problèmes unilatéraux doivent alors conduire à un vote en faveur de Θ_1^- plus faible que celui en faveur de $\Theta_1^- \cup \Theta_0$. Nous dirons qu'ils sont compatibles, ils permettent alors d'induire une probabilité sur la tribu engendrée par les trois hypothèses. Les trois problèmes de décision considérés ne sont généralement pas les seuls envisageables dans une application donnée. Il est en effet rare que les bornes θ_1 et θ_2 ne soient pas sujettes à discussion. Aussi nous avons étendu la notion de votes compatibles à un ensemble quelconque d'hypothèses unilatérales.

Définition 4.1.
Soient $(\{\Theta_1^f, \Theta_0^f\})_{f\in\mathcal{F}}$ *une famille d'hypothèses unilatérales et* Q *une application de* $\mathbb{R} \times \{\Theta_1^f\}_{f\in\mathcal{F}}$ *dans* $[0,1]$ *telle que* $Q^x(\Theta_1^f)$ *représente la valeur en* $d = 1$ *d'un vote d'experts du choix entre* Θ_1^f *et* Θ_0^f *lorsqu'on réalise* x.
Nous dirons que Q *définit des votes compatibles lorsqu'elle se prolonge, pour presque tout* x, *en une probabilité unique* Q^x *sur la tribu engendrée par* $\{\Theta_1^f\}_{f\in\mathcal{F}}$.

Dans le cadre de cet article nous n'étudierons la compatibilité des votes que pour les votes neutres sélectionnés au paragraphe précédent (l'introduction d'informations a priori est traitée dans Morel (1997)). Pour les hypothèses unilatérales $\{]-\infty, \theta],]\theta, +\infty[\}$ le vote neutre est celui défini par le point frontière θ. Nous allons donc considérer l'application :
$Q^x(]-\infty, \theta]) = 1 - F(\theta, x)$, $F(\theta, .)$ étant la fonction de répartition de $\mathbb{P}_\theta$.

Proposition 4.2. *Sur un intervalle* $I \subseteq \mathbb{R}$ *muni de la mesure de Lebesgue, considérons une famille* $\{p_\theta\}_{\theta\in\Theta}$ *de densités strictement positives.* Θ *étant*

un intervalle de $\mathbb{R}$, cette famille est supposée continue en θ et à rapport de vraisemblance strictement monotone.
Pour tout $\theta \in \Theta - \{sup\Theta\}$ considérons le vote neutre défini par les hypothèses unilatérales $\Theta_1^\theta =]-\infty, \theta] \cap \Theta$ et $\Theta_0^\theta =]\theta, +\infty[\cap\Theta$. Ces votes sont compatibles si et seulement si, pour presque tout $x \in I$ on a :
1) $lim_{\theta \to sup\Theta} F(\theta, x) = 0$ si $sup\Theta \notin \Theta$
2) $lim_{\theta \to inf\Theta} F(\theta, x) = 1$ si $inf\Theta \notin \Theta$.

Démonstration.

Les votes neutres définissent une application Q de $\mathbb{R} \times \{\Theta_1^\theta\}_{\theta \in \Theta - sup\Theta}$ dans $[0,1]$: $Q^x(\Theta_1^\theta) = 1 - F(\theta, x)$. Q donne des votes compatibles si elle se prolonge, pour presque tout x, en une probabilité unique sur les Boréliens de l'intervalle Θ.
Pour une réalisation x donnée, l'application Q^x est prolongeable si elle vérifie les propriétés caractéristiques d'une fonction de répartition.
a) La croissance en θ de $Q^x(\Theta_1^\theta) = 1 - F(\theta, x)$ est une conséquence directe de la décroissance de $F(., x)$ sur une famille à rapport de vraisemblance monotone.
b) La continuité à droite en θ de $Q^x(\Theta_1^\theta)$ est équivalente à celle de $F(\theta, x)$. La famille des densités p_θ étant continue, elle l'est aussi dans L_1, ce qui implique la continuité de $F(., x)$.
c) Il reste à vérifier les conditions aux bornes, c'est-à-dire la convergence de $Q^x(\Theta_1^\theta)$ vers 1 (resp. 0) lorsque Θ_1^θ croît vers Θ (resp. décroît vers $\emptyset$). Une telle situation n'est possible que si $sup\Theta \notin \Theta$ (resp. $inf\Theta \notin \Theta$). Les deux conditions nécessaires et suffisantes de la proposition en découlent immédiatement. □

Un ensemble de votes compatibles sur les hypothèses unilatérales permet de prolonger de façon cohérente ces expertises à des hypothèses non expertisables. Sous les conditions de la proposition précédente, on obtient pour les hypothèses bilatérales $\Theta_1 =]-\infty, \theta_1[\cup]\theta_2, +\infty[$ et $\Theta_0 = [\theta_1, \theta_2]$ les votes suivants : $Q^x(\Theta_1) = 1 - F(\theta_1, x) + F(\theta_2, x)$ et $Q^x(\Theta_0) = F(\theta_1, x) - F(\theta_2, x)$. Ces votes ne sont pas les p-values données par les tests classiques. Ils nous semblent préférables lorsque la structure d'ordre sur Θ reste primordiale pour l'interprétation, malgré la cassure introduite par l'hypothèse Θ_1 qui regroupe $]-\infty, \theta_1[$ et $]\theta_2, +\infty[$. On évite ainsi les incohérences des p-values relevées par Schervish (1996). Les votes obtenus sans information a priori définissent sur Θ la distribution fiduciaire de Fisher.

Exemple 4.3.

Considérons le cas traité par Schervish (1996), celui de la famille des lois normales $\{N(\theta, 1)\}_{\theta \in \mathbb{R}}$. La proposition 4.2 s'applique : $F(\theta, x) = \Phi(x - \theta)$, Φ étant la fonction de répartition de la loi $N(0,1)$. La probabilité Q^x sur Θ est la loi $N(x, 1)$. C'est la loi a posteriori correspondant à la mesure de Lebesgue, loi a priori impropre unanimement considérée comme non informative.

La probabilité inductive Q^x obtenue à partir de votes neutres compatibles correspond souvent à la loi a posteriori d'une loi a priori qui peut être

considérée comme non informative [Morel(1997)]. Dans le cadre de ce paragraphe citons la famille des lois normales $\{N(m,\theta^2)\}_{\theta\in\mathbb{R}_*^+}$ et celle des lois gamma $\{\gamma(p,\theta)\}_{\theta\in\mathbb{R}_*^+}$ (voir l'exemple 5.1).

4.2. P-values des tests sans biais comme votes d'experts

Nous considérons maintenant le cas où les deux parties $\Theta_1^- =]-\infty,\theta_1[$ et $\Theta_1^+ =]\theta_2,+\infty[$ de Θ_1 sont équivalentes pour l'interprétation. Celle-ci est censée évoluer de la même manière de $-\infty$ à θ_1 et de $+\infty$ à θ_2.

Dans l'exemple simple de la famille des lois normales $\{N(\theta,1)\}_{\theta\in\mathbb{R}}$, le problème de décision peut être considéré comme invariant par symétrie autour de $c = (\theta_1+\theta_2)/2$. Cette notion sert classiquement à diminuer l'ensemble des règles possibles. Ici nous devons au contraire augmenter le nombre d'experts. Pour cela nous allons restreindre la tribu du modèle aux événements invariants, les boréliens symétriques par rapport à c. Cette restriction sur les événements impose des contraintes moins fortes pour le label expert. Elle permet d'avoir d'autres experts que les experts triviaux du modèle de départ. En fait on travaille sur le modèle associé à la statistique $W(x) = (x-c)^2$. C'est exactement ce que l'on fait, quand pour comparer deux moyennes, on utilise la statistique $W = T^2$ de Fisher au lieu de la statistique T de Student. Dans le nouveau modèle θ et $2c-\theta$ définissent la même probabilité, la loi de khi-deux décentrée à un degré de liberté et de paramètre d'excentricité $\lambda^2 = (\theta-c)^2$. On peut prendre comme ensemble des paramètres $\Lambda = \mathbb{R}^+$, le choix entre Θ_0 et Θ_1 devient un choix entre $[0,(\theta_2-\theta_1)/2]$ et $](\theta_2-\theta_1)/2,+\infty[$. Ce sont des hypothèses unilatérales dans un modèle à rapport de vraisemblance monotone [Karlin (1955)]. La proposition 3.1, donne les experts de ce nouveau problème de décision paramétré par λ. Avec le paramètre θ on a le résultat suivant.

Corollaire 4.4.
Considérons le modèle statistique $(\mathbb{R},\mathcal{C},\{N(\theta,1)\}_{\theta\in\mathbb{R}})$*,* $\mathcal{C}$ *étant la tribu des boréliens symétriques par rapport à* $c\in\mathbb{R}$*. Posons* $\theta_1 = c-\lambda_1$ *et* $\theta_2 = c+\lambda_1$ *avec* $\lambda_1 \geq 0$*. Les experts du choix entre les deux hypothèses* $\Theta_1 =]-\infty,\theta_1[\cup]\theta_2,+\infty[$ *et* $\Theta_0 = [\theta_1,\theta_2]$ *sont les règles de décision à valeurs dans* $\{0,1\}$*, presque sûrement de la forme :*
$g_t(x) = \mathbb{1}_{[t,+\infty[}(|x-c|)$ *avec* $t\in\overline{\mathbb{R}^+}$*.*

Les hypothèses unilatérales $\Lambda_1 = [0,\lambda_1]$ et $\Lambda_0 =]\lambda_1,+\infty[$ sont équivalentes aux hypothèses Θ_0 et Θ_1. Comme au paragraphe 3 on peut définir le vote au point frontière λ_1 ; lorsqu'on réalise x, le vote en faveur de Λ_0, c'est-à-dire Θ_1, est alors donné par la valeur en $W(x) = (x-c)^2$ de la fonction de répartition d'une loi de khi-deux décentrée à un degré de liberté et de paramètre d'excentricité $\lambda_1^2 = (\theta_1-c)^2$. Notons $Q^x(\Theta_1)$ ce vote, c'est la probabilité de l'intervalle $]-|x-c|,|x-c|[$ pour la loi $N(\theta_1-c,1)$.
$Q^x(\Theta_1) = \Phi(|x-c|-\theta_1+c) - \Phi(-|x-c|-\theta_1+c)$ est donc la p-value du test de $H_0' : \theta\in\Theta_1$ contre $H_1' : \theta\in\Theta_0$. Le vote $Q^x(\Theta_0)$ en faveur de Θ_0 est bien sûr la p-value du test sans biais de $H_0 : \theta\in\Theta_0$ contre $H_1 : \theta\in\Theta_1$. Ceci reste valable pour le cas particulier $\theta_1 = \theta_2 = c$ qui correspond à l'hypothèse

$\Theta_0 = [c]$. La probabilité sur $\Lambda = \mathbb{R}^+$ donnée par la proposition 4.2 possède une masse en $\{c\}$ qui est la p-value du test de $H_0 : \theta = c$ contre $H_1 : \theta \neq c$. Sur cet exemple, nous venons de retrouver comme votes d'experts les p-values des tests bilatéraux classiques. Ceci n'a été possible qu'en supposant que l'interprétation du paramètre présente une symétrie par rapport au centre de l'intervalle $[\theta_1, \theta_2]$. Ce n'est pas le cas lorsque l'interprétation reste liée à l'ordre sur Θ, il est alors préférable d'utiliser les votes compatibles du paragraphe précédent.
Ce que nous venons de faire sur l'exemple simple des lois $\{N(\theta, 1)\}$ peut se généraliser aux autres modèles exponentiels. La notion de sans biais pour le test de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ et celle de seuil pour le test de $H'_0 : \theta \in \Theta_1$ contre $H'_1 : \theta \in \Theta_0$ imposent des règles de décision dont la région de rejet C vérifie $\mathbb{P}_{\theta_1}(C) = \mathbb{P}_{\theta_2}(C)$. Cette restriction sur les règles possibles est une manière d'imposer un traitement "symétrique" des deux parties Θ_1^- et Θ_1^+ de l'hypothèse Θ_1. Pour retrouver les p-values de ces tests comme votes d'experts, il faut rendre le problème de décision expertisable en enlevant des contraintes dans la définition 2.1. La recherche d'une symétrie de traitement des sous hypothèses Θ_1^- et Θ_1^+ se traduira par la restriction aux événements C vérifiant $\mathbb{P}_{\theta_1}(C) = \mathbb{P}_{\theta_2}(C)$.

5. Modèles avec paramètre fantôme

L'espace des paramètres est un espace produit $\Theta \times \Upsilon$, $\theta \in \Theta \subseteq \mathbb{R}$ est le paramètre d'intérêt et $\upsilon \in \Upsilon$ le paramètre fantôme. On considère le problème du choix entre $\Theta_1 \times \Upsilon$ et $\Theta_0 \times \Upsilon$. Pour chaque valeur possible υ du paramètre fantôme le problème de décision se réduit au type de ceux que nous avons étudiés précédemment. Une expertise, si elle existe, donnera des votes, en faveur de chacune des hypothèses Θ_1 et Θ_0, indexés par $\upsilon \in \Upsilon$. Ces résultats conditionnels au paramètre fantôme peuvent être résumés en utilisant une probabilité sur Υ. Pour définir cette probabilité nous chercherons à utiliser des votes compatibles.

Nous allons donner des exemples dans le cas simple mais courant où le problème de décision repose sur deux statistiques indépendantes réelles (T, U), qui définissent chacune un modèle à rapport de vraisemblance monotone par rapport à θ respectivement υ, la loi de T pouvant dépendre de υ (d'autres cas sont traités dans Morel (1997)).

Exemple 5.1.
Commençons par étudier le cas classique d'un n-échantillon d'une loi $N(\theta, \upsilon = \sigma^2)$, $n \geq 2$, $\theta \in \mathbb{R}$ et $\upsilon \in \mathbb{R}_*^+$. La moyenne et la variance empiriques, $\overline{X}$ et $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, sont exhaustives et indépendantes. La statistique $T = \overline{X}$ est de loi $N(\theta, \upsilon/n)$, elle définit bien un modèle à rapport de vraisemblance monotone par rapport à θ. Quant à la statistique $U = (n-1)S^2$ elle suit une loi de khi-deux à $(n-1)$ degrés de liberté et ayant pour paramètre d'échelle υ, c'est-à-dire une loi gamma $\gamma(\frac{n-1}{2}, 2\upsilon)$. Elle définit un modèle statistique à rapport de vraisemblance strictement monotone par rapport à υ [Karlin (1955)].

Considérons les hypothèses unilatérales : $\Theta_1^\theta =]-\infty, \theta]$ et $\Theta_0^\theta =]\theta, +\infty[$. Pour toute valeur fixée υ du paramètre fantôme, nous savons expertiser ce problème de décision. Il dépend uniquement de la statistique T. La proposition 4.2 s'applique et pour une réalisation $t = \overline{x}$ les votes neutres compatibles définissent sur $\Theta = \mathbb{R}$ la loi $N(\overline{x}, \upsilon/n)$.
La statistique U indépendante de T va nous permettre de probabiliser $\Upsilon = \mathbb{R}_*^+$. En effet elle définit sur $\mathbb{R}^+$ muni de la mesure de Lebesgue la famille de densités $\{\frac{1}{\Gamma(p)2\upsilon}(\frac{u}{2\upsilon})^{p-1} exp(-\frac{u}{2\upsilon})\}_{\upsilon \in \mathbb{R}_*^+}$, avec $p = \frac{n-1}{2}$. Ces densités sont nulles uniquement en $u = 0$ qui est négligeable et la fonction de répartition est donnée par $F(\upsilon, u) = \int_0^{\frac{u}{2\upsilon}} \frac{1}{\Gamma(p)} y^{p-1} e^{-y}\, dy$, on peut donc appliquer la proposition 4.2. Elle nous donne les votes compatibles suivant :
$Q^u(]0, \upsilon]) = 1 - F(\upsilon, u) = \int_0^\upsilon \frac{2}{\Gamma(p)u}(\frac{u}{2\lambda})^{p+1} exp(-\frac{u}{2\lambda})\, d\lambda \ (u > 0)$.
Leur prolongement est donc la loi inverse d'une loi $\gamma(p, \frac{2}{u})$, notée $Inv\gamma(p, \frac{2}{u})$. C'est la loi a posteriori pour la loi a priori impropre et non informative de densité $\frac{1}{\upsilon}$ [Berger (1985)], mais aussi la distribution fiduciaire [Fisher (1935)].
Avec cette probabilité inductive nous devons maintenant faire la moyenne des votes obtenus conditionnellement à υ. Si on réalise $t = \overline{x}$ et $u = (n-1)s^2$, le vote en faveur de $\Theta_1^\theta =]-\infty, \theta]$, noté $Q^{(t,u)}(]-\infty, \theta])$, s'obtient en faisant la moyenne de $\Phi(\frac{\theta - t}{\sqrt{\upsilon/n}})$, Φ étant la fonction de répartition de la loi $N(0, 1)$.

$$
\begin{aligned}
Q^{(t,u)}(]-\infty, \theta]) &= \int \Phi((\theta - t)/\sqrt{\upsilon/n})\, d\, Inv\gamma(p, \tfrac{2}{u})(\upsilon) \\
&= \int_{\mathbb{R}^+} \Phi(\sqrt{n}(\theta - t)\sqrt{\lambda})\, d\,\gamma(p, \tfrac{2}{u})(\lambda) \\
&= \int_{\mathbb{R}^+} \Phi(\sqrt{n}(\theta - t)\sqrt{\lambda}) \frac{u}{2\Gamma(p)}(\tfrac{u}{2}\lambda)^{p-1}\, e^{-\frac{u}{2}\lambda}\, d\lambda \\
&= \int_{\mathbb{R}^+} [\int_{-\infty}^{\sqrt{n}(\theta - t)} \frac{\sqrt{\lambda}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}\lambda}\, dx] \frac{u}{2\Gamma(p)}(\tfrac{u}{2}\lambda)^{p-1}\, e^{-\frac{u}{2}\lambda}\, d\lambda \\
&= \int_{\mathbb{R}^+} [\int_{-\infty}^{\frac{\sqrt{n}(\theta - t)}{\sqrt{s^2}}} \frac{\sqrt{s^2\lambda}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}s^2\lambda}\, dy] \frac{u}{2\Gamma(p)}(\tfrac{u}{2}\lambda)^{p-1}\, e^{-\frac{u}{2}\lambda}\, d\lambda \\
&= \int_{-\infty}^{\frac{\sqrt{n}(\theta - t)}{\sqrt{s^2}}} [\int_0^{+\infty} \frac{\sqrt{\nu}}{\sqrt{2\pi}} e^{-\frac{y^2}{2}\nu} \times \frac{(\frac{n-1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \nu^{\frac{n-1}{2}-1}\, e^{-\frac{n-1}{2}\nu}\, d\nu]\, dy
\end{aligned}
$$

L'intégrale entre crochets donne la densité d'une loi de Student à $(n-1)$ degrés de liberté comme mélange des lois $N(0, \frac{1}{\nu})$ par la loi $\gamma(\frac{n-1}{2}, \frac{2}{n-1})$ [Dickey (1968)]. Le vote $Q^{(t,u)}(]-\infty, \theta])$ est donc égal à la valeur de la fonction de répartition d'un Student $(n-1)$ en $\frac{\sqrt{n}(\theta - t)}{\sqrt{s^2}} = -\sqrt{n}\frac{(\overline{x} - \theta)}{\sqrt{s^2}}$. C'est le seuil minimum de rejet du test de Student de $\Theta_1^\theta =]-\infty, \theta]$ contre $\Theta_0^\theta =]\theta, +\infty[$. Bien sûr $Q^{(t,u)}(\Theta_0^\theta) = 1 - Q^{(t,u)}(\Theta_1^\theta)$ est le seuil minimum de rejet du test de Student de Θ_0^θ contre Θ_1^θ.
Lorsque θ parcourt $\mathbb{R}$, les votes précédents sont compatibles. Ils définissent, sur l'espace $\mathbb{R}$ du paramètre θ, une probabilité qui est une loi de Student à $(n-1)$ degrés de liberté, de moyenne $\overline{x}$ et de paramètre d'échelle $\frac{s^2}{n}$ [Dickey (1968)], c'est la distribution fiduciaire donnée par Fisher (1935).

Considérons maintenant le cas des hypothèses bilatérales
$\Theta_1 =]-\infty, \theta_1[\cup]\theta_2, +\infty[$ et $\Theta_0 = [\theta_1, \theta_2]$. Si la structure d'ordre sur $\Theta = \mathbb{R}$ est primordiale pour l'interprétation, il nous semble normal de prendre un vote cohérent par rapport aux votes correspondant aux hypothèses unilatérales. Avec les votes neutres on obtient la probabilité inductive de Θ_1 et Θ_0 donnée par la loi précédente. Ce ne sont pas les p-values des tests de Student bilatéraux. Comme dans le cas des lois $N(\theta, 1)$ on retrouve ces

p-values en utilisant la symétrie par rapport à $c = (\theta_1 + \theta_2)/2$ du problème de décision pour le rendre expertisable. C'est en fait travailler sur θ à partir des statistiques indépendantes $(\overline{X} - c)^2$ et S^2. Nous allons le faire dans un cadre plus général, celui de l'analyse de variance.

Exemple 5.2.

Reprenons l'exemple 3.3 de l'analyse de variance à effets fixes. Nous avons travaillé sur le paramètre de non centralité λ d'une loi de Fisher décentrée. λ dépend du paramètre fantôme σ^2, la variance inconnue. On a $\lambda^2 = \theta/\sigma^2$, le paramètre $\sqrt{\theta}$ est en fait généralement celui qui intéresse l'utilisateur, son expression λ en unité d'écart-type est un produit de la statistique. D'ailleurs pour les tests courants de comparaison de λ à 0, on exprime $H_0 : \lambda = 0$ par $\theta = 0$. Le paramètre λ n'apparaît que dans la puissance de ces tests.

La statistique de Fisher décentrée W que nous avons utilisée précédemment est construite à partir de deux statistiques indépendantes T et U ; T/σ^2 suit une loi de khi-deux décentrée à k degrés de liberté et de paramètre de non centralité θ ; U/σ^2 suit une loi de khi-deux à l degrés de liberté [Scheffé (1970)]. La réduction du problème de base à l'étude du modèle engendré par (T, U) se fait en imposant des propriétés d'invariance. Le modèle image s'écrit $((\mathbb{R}^+)^2, \mathcal{B}^2, (P^T_{(\theta,\sigma^2)} \otimes P^U_{\sigma^2})_{(\theta,\sigma^2)\in\mathbb{R}^+\times\mathbb{R}^+_*})$, $P^T_{(\theta,\sigma^2)}$ est une loi $\gamma(\frac{k}{2}, 2\sigma^2)$ décentrée de $\theta/(2\sigma^2)$ et $P^U_{\sigma^2}$ une loi $\gamma(\frac{l}{2}, 2\sigma^2)$.

Plus généralement nous allons étudier les modèles statistiques de la forme : $((\mathbb{R}^+)^2, \mathcal{B}^2, (P^T_{(\theta,\upsilon)} \otimes P^U_{\upsilon})_{(\theta,\upsilon)\in\mathbb{R}^+\times\mathbb{R}^+_*})$ où $P^T_{(\theta,\upsilon)}$ est une loi $\gamma(p, \upsilon)$ décentrée de θ/υ et P^U_{υ} une loi $\gamma(q, \upsilon)$. Nous allons procéder comme dans l'exemple précédent.

Considérons les hypothèses unilatérales : $\Theta^\theta_1 = [0, \theta]$ et $\Theta^\theta_0 =]\theta, +\infty[$ $(\theta \geq 0)$.

Pour υ fixé le problème de décision ne dépend que de la statistique T donc des lois $\{P^T_{(\theta,\upsilon)}\}_{\theta\in\mathbb{R}^+}$ ou de la famille des densités $\{\int_{\mathbb{N}} \frac{1}{\Gamma(p+m)\upsilon} (\frac{t}{\upsilon})^{p+m-1} exp(-\frac{t}{\upsilon})\, d\mathcal{P}_{\frac{\theta}{\upsilon}}(m)\}_{\theta\in\mathbb{R}^+}$, $\mathcal{P}_{\frac{\theta}{\upsilon}}$ étant une loi de poisson de paramètre $\frac{\theta}{\upsilon} \geq 0$ [Barra (1971)]. On est ramené au choix entre deux hypothèses unilatérales dans un modèle à rapport de vraisemblance strictement monotone [Karlin (1955)]. Si l'on ne veut ou peut pas faire intervenir une information a priori on utilise le vote neutre, celui au point frontière θ. Pour la réalisation t la fréquence des experts en faveur de $\Theta^\theta_1 = [0, \theta]$ est alors égale à $1 - F_\upsilon(\theta, t)$, $F_\upsilon(\theta, t)$ étant la valeur en t de la fonction de répartition de $P^T_{(\theta,\upsilon)}$.

$F_\upsilon(\theta, t) = \int_{\mathbb{N}} \Gamma(p + m, \upsilon, t)\, d\mathcal{P}_{\frac{\theta}{\upsilon}}(m)$, le terme $\Gamma(p + m, \upsilon, t)$ désignant la valeur de la fonction de répartition d'une loi $\gamma(p + m, \upsilon)$ en t :

$\Gamma(p + m, \upsilon, t) = \int_0^t \frac{1}{\Gamma(p+m)\upsilon} (\frac{x}{\upsilon})^{p+m-1} exp(-\frac{x}{\upsilon})\, dx = \Gamma(p + m, 1, \frac{t}{\upsilon})$.

Nous allons maintenant probabiliser $\Upsilon = \mathbb{R}^+_*$ en utilisant la statistique U qui est de loi $\gamma(q, \upsilon)$. Nous avons étudié ce type de modèle dans l'exemple précédent. Les votes neutres pour les hypothèses unilatérales sont compatibles. Ils se prolongent en une probabilité sur les boréliens de $\Upsilon = \mathbb{R}^+_*$ qui est l'inverse d'une loi $\gamma(q, \frac{1}{u})$ $(u > 0)$.

Pour toute réalisation (t, u), la moyenne $Q^{(t,u)}([0, \theta])$ des votes en faveur de $[0, \theta]$ est alors égale à :

$$
\begin{aligned}
Q^{(t,u)}([0,\theta]) &= 1-\int_{\mathbb{R}_*^+} F_\upsilon(\theta,t)\frac{1}{\Gamma(q)u}\,(\tfrac{u}{\upsilon})^{q+1}\,exp(-\tfrac{u}{\upsilon})\,d\upsilon \\
&= 1-\int_{\mathbb{R}_*^+}\int_{\mathbb{N}}[\int_0^t \frac{1}{\Gamma(p+m)\upsilon}\,(\tfrac{x}{\upsilon})^{p+m-1}\,exp(-\tfrac{x}{\upsilon})\,dx]\,d\mathcal{P}_{\frac{\theta}{\upsilon}}(m)\times \\
&\qquad\qquad\qquad\qquad \frac{1}{\Gamma(q)u}\,(\tfrac{u}{\upsilon})^{q+1}\,exp(-\tfrac{u}{\upsilon})\,d\upsilon \\
&= 1-\sum_{m=0}^{+\infty}\int_0^t[\int_{\mathbb{R}_*^+}\frac{exp(-\frac{\theta}{\upsilon})\,(\frac{\theta}{\upsilon})^m}{m!\Gamma(p+m)\upsilon}\,(\tfrac{x}{\upsilon})^{p+m-1}\,exp(-\tfrac{x}{\upsilon})\,\frac{1}{\Gamma(q)u}\times \\
&\qquad\qquad\qquad\qquad (\tfrac{u}{\upsilon})^{q+1}\,exp(-\tfrac{u}{\upsilon})\,d\upsilon]\,dx \\
&= 1-\sum_{m=0}^{+\infty}\int_0^t\int_{\mathbb{R}_*^+}\frac{\theta^m\,u^q\,x^{p+m-1}}{m!\Gamma(p+m)\Gamma(q)}(\tfrac{1}{\upsilon})^{p+q+2m+1}exp(-\tfrac{\theta+u+x}{\upsilon})d\upsilon dx \\
&= 1-\sum_{m=0}^{+\infty}\int_0^t\frac{\Gamma(p+q+2m)}{m!\Gamma(p+m)\Gamma(q)}\,\frac{\theta^m\,u^q\,x^{p+m-1}}{(\theta+u+x)^{p+q+2m}}\,dx \\
&= 1-\sum_{m=0}^{+\infty}\int_0^{\frac{t}{\theta+u}}\frac{\Gamma(p+q+2m)}{m!\Gamma(p+m)\Gamma(q)}\,\frac{\theta^m\,u^q}{(\theta+u)^{q+m}}\,\frac{y^{p+m-1}}{(1+y)^{p+q+2m}}\,dy \\
&= 1-\sum_{m=0}^{+\infty}\frac{\Gamma(q+m)}{m!\Gamma(q)}\,\frac{\theta^m\,u^q}{(\theta+u)^{q+m}}\,F(p+m,q+m,\tfrac{t}{\theta+u})
\end{aligned}
$$

$F(p+m,q+m,\frac{t}{\theta+u})$ étant la valeur en $\frac{t}{\theta+u}$ de la fonction de répartition d'une loi $\beta(p+m,q+m)$ sur $\mathbb{R}^+$. Pour $\theta=0$ on obtient une masse $Q^{(t,u)}([0]) = 1-F(p,q,\frac{t}{u})$. Dans le cas de l'analyse de variance, $p=\frac{k}{2}$ et $q=\frac{l}{2}$, cette masse est égale à la probabilité qu'une loi de Fisher, de degrés de liberté k et l, soit supérieure à $\frac{t/k}{u/l}$. C'est le seuil minimum de rejet du test de $H_0:\theta=0$ contre $H_1:\theta>0$. Nous avions déjà obtenu ce type de résultat pour le paramètre $\frac{\theta}{\sigma^2}$ dans l'exemple 3.3.

Ce qui précède nous permet de définir une probabilité inductive $Q^{(t,u)}$ sur $\Theta=\mathbb{R}^+$ puisque pour tout $\upsilon\in\Upsilon$ la proposition 4.2 s'applique (la fonction de répartition $F_\upsilon(\theta,t)$ tend bien vers 0 lorsque θ tend vers $+\infty$ car $\Gamma(p+m,\upsilon,t)=\Gamma(p+m,1,\frac{t}{\upsilon})$ tend vers 0 lorsque m tend vers $+\infty$). On peut ainsi traiter d'autres hypothèses que les hypothèses unilatérales.

Comme en 3.3 les expertises obtenues dans cet exemple complètent l'information donnée par les résultats des tests classiques, mais cette fois directement sur la valeur du paramètre étudié et non pas sur sa valeur exprimée en écart-type. Le seuil minimum de rejet $\alpha_m(\frac{t}{u})$ du test de $H_0:\theta=0$ contre $H_1:\theta>0$ est vu comme le vote $Q^{(t,u)}([0])$ des experts en faveur de l'hypothèse $H_0:\theta=0$. Dans le cas du non rejet de H_0, c'est-à-dire lorsque $\alpha_m(\frac{t}{u})$ est supérieur au seuil choisi, on peut préciser cette réponse en regardant si $Q^{(t,u)}([0,\theta])$ se rapproche rapidement de 1 lorsque θ croît. Le cas du rejet de H_0 pose problème lorsque l'hypothèse $H_0:\theta=0$ est une idéalisation de l'hypothèse réelle à tester. Bien souvent l'utilisateur se demande si θ est petit et non pas si θ est nul. Une interprétation trop rapide du rejet peut conduire à considérer que θ est notable alors qu'il est négligeable. Une analyse de la croissance de $Q^{(t,u)}([0,\theta])$ lorsque θ s'éloigne de 0 permet d'éviter facilement ce piège. Il est cependant plus satisfaisant d'essayer de traduire l'hypothèse "θ est petit" par $\theta\in[0,\theta_0]$ et de porter un jugement à partir de $Q^{(t,u)}([0,\theta_0])$.

La prise en compte des votes $Q^{(t,u)}([0,\theta])$ autour de l'hypothèse nulle $\theta=0$ donne une manière simple et parlante de compléter l'information donnée par les tests classiques. Elle est moins complexe à analyser que la notion de puissance et de plus elle ne dépend pas de l'écart-type inconnu σ.

6. Conclusion

Dans des modèles à rapport de vraisemblance monotone sur $\mathbb{R}$ nous avons vu que la p-value associée à un test unilatéral est généralement considérée comme un indice de la confiance qu'on accorde à l'hypothèse H_0. Avec notre approche cet indice s'interprète comme le résultat d'un vote des experts en faveur de H_0, ce vote étant défini sans information a priori. La notion d'expertises compatibles pour l'ensemble des hypothèses unilatérales nous a conduit à prolonger ces votes en une probabilité inductive Q^x sur l'espace des paramètres, x étant la réalisation obtenue. On retrouve ainsi la distribution fiduciaire de Fisher. La probabilité Q^x est souvent une loi a posteriori associée à une mesure a priori qui peut être considérée comme non informative dans le cadre bayésien.

Prendre en compte l'ensemble des hypothèses unilatérales n'a de sens que si l'ordre sur l'espace des paramètres est totalement structurant pour l'interprétation. Dans ce cas nous avons proposé, pour des raisons de cohérence, de prendre comme indice de confiance d'une hypothèse de la forme $\theta \in [\theta_1, \theta_2]$ la valeur sur cet intervalle de la probabilité induite Q^x. C'est le prolongement aux hypothèses bilatérales du vote des experts unilatéraux. Ce n'est pas la p-value du test bilatéral associé à $H_0 : \theta \in [\theta_1, \theta_2]$. Cette manière de faire permet de prendre en compte les trois hypothèses : $]-\infty, \theta_1[$, $[\theta_1, \theta_2]$, $]\theta_2, +\infty[$, et d'avoir des interprétations différentes pour chacune d'elles. Les tests bilatéraux ne sont pas construits pour cela.
Les problèmes soulevés par les p-values des deux tests bilatéraux définis à partir de l'hypothèse $\theta \in [\theta_1, \theta_2]$ proviennent de l'équivalence d'interprétation entre $\theta < \theta_1$ et $\theta > \theta_2$ que ces tests supposent. Dans l'exemple des lois $N(\theta, 1)$ ou $N(\theta, \sigma^2)$, nous avons retrouvé leurs p-values comme votes d'experts en ne considérant que les événements symétriques par rapport à $c = (\theta_1 + \theta_2)/2$. Ces p-values nous semblent recommandables uniquement dans le cas où l'interprétation est la même que θ soit inférieur à θ_1 ou supérieur à θ_2. La structure du problème de décision change. L'ordre sur Θ n'est plus alors totalement structurant pour l'interprétation, c'est la distance à c qui devient primordiale. Les hypothèses légitimes sont invariantes pour cette distance, elles sont symétriques par rapport à c et leurs p-values sont cohérentes. On peut généraliser cette manière de faire aux modèles exponentiels réels. La notion de cohérence des votes nous impose de traiter globalement un ensemble de problèmes de décision, sur un modèle statistique donné.
Le choix entre les deux types de traitement des hypothèses bilatérales dépend de la manière dont l'utilisateur structure son interprétation des différentes valeurs du paramètre. C'est la structuration par l'ordre sur $\Theta \subseteq \mathbb{R}$ qui est la plus courante. Dans ce cas les p-values des tests bilatéraux ne devraient pas être utilisées.

Au départ la notion d'expertise n'a pas été construite pour traiter le problème des p-values mais pour essayer de chercher une aide à la décision sous la forme d'un indice de confiance pour chacune des hypothèses. Ce type d'inférence a été défendu dans de nombreux travaux (Kroese, van der Meulen, Poortema et Schaafsma (1995)). Il est particulièrement adapté

aux applications qui ne réclament pas un choix définitif entre les deux hypothèses. Les solutions les plus courantes se font dans le cadre bayésien ou celui de l'estimation. Nous avons essayé de nous passer du choix souvent difficile d'une loi a priori ou d'une fonction de perte. Le cas particulier des votes neutres, compatibles sur les hypothèses unilatérales, donne des résultats qui sont à rapprocher de ceux de l'analyse fiduciaire de Fisher. Utilisée sans information a priori, la notion d'expertise s'inscrit dans le champ de questionnement ouvert par ce que l'on peut appeler le compromis de Fisher-Neyman-Wald.

Bibliographie

Barra J. R. (1971), *Notions fondamentales de statistique mathématique*, Dunod, Paris.

Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis* (second edition), Springer-Verlag, New York.

Berger, J. O. et Sellke, T. (1987), Testing a point null hypothesis: the irreconcilability of p-values and evidence, *J. Amer. Statist. Assoc.* **82** 112–122.

Buehler, R. J. (1980), Fiducial inference. *R. A. Fisher : an appreciation*, Springer-Verlag, New York, Lecture Notes in Statistics 109-118.

Casella, G. et Berger, R. L. (1987), Reconciling Bayesian and frequentist evidence in the one-sided testing problem, *J. Amer. Statist. Assoc.* **82** 106–111.

Dickey, J. M. (1968), Three multidimensional-integral identities with bayesian applications, *Ann. Math. Statist.* **39** 1615–1628.

Fisher, R. A. (1935), The fiducial argument in statistical inference, *Annals of Eugenics* **6** 391–398.

Gabriel, K. R. (1969), Simultaneous test procedures — some theory of multiple comparisons, *Ann. Math. Statist.* **40** 224–250.

Hung, H. M. J., O'Neill, R. T., Bauer, P. et Köhne, K. (1997), The behavior of the p-value when the alternative hypothesis is true, *Biometrics* **53** 11–22.

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., Farrell, R. H. (1992), Estimation of accuracy in testing, *Ann. Statist.* **20** 490–509.

Karlin, S. (1955), Decision theory for Pólya type distributions. Case of two actions, I, *Proc. Third Berkeley Symposium on Math. Statist. and Prob.* (Univ. of Calif. Press) **1** 115–128.

Kroese, A. H., van der Meulen, E. A., Poortema, K., Schaafsma, W. (1995), Distributional inference, *Statistica Neerlandica* **49** 63–82.

Lehmann, E. L. (1986), *Testing statistical hypotheses* (second edition), Wiley, New York.

van der Meulen, E. A., Schaafsma, W. (1993), Assessing weights of evidence for discussing classical statistical hypotheses, *Statistics & Decisions* **11** 201–220.

Monfort, A. (1982), *Cours de statistique mathématique*, Economica, Paris.

Morel, G. (1997), Expertises : procédures statistiques d'aide à la décision, *Pré-publication*, 182 p., LAST-Université de Tours.

Morel, G. (1998), Probabiliser l'espace des décisions, *Pub. Sém. 97 : Bru Huber Prum,* Paris V, 71–98.

Robert, C. (1992), *L'analyse statistique bayésienne*, Economica, Paris.

Salomé, D. (1998), *Statistical inference via fiducial methods*, PhD thesis, Rijksuniversiteit Groningen.

Schaafsma, W., Tolboom, J. et van der Meulen, B. (1989), Discussing truth and falsity by computing a Q-value, in *Statistical Data Analysis and Inference* (Y. Dodge, ed.), North-Holland, Amsterdam, 85–100.

Scheffé H. (1970), *The analysis of variance* (sixième édition), Wiley, New York.

Schervish, M. J. (1996), P-values : what they are and what they are not, *Amer. Statist.* **50** 203–206.

Thompson, P. (1996), Improving the admissibility screen : evaluating test statistics on the basis of p-values rather than power, *Comm. Statist. Theory Methods* **25** 537–553.